

NONVOLATILE MEMORY FABRICATION METHODS COMPRISING LATERAL
RECESSING OF DIELECTRIC SIDEWALLS
AT SUBSTRATE ISOLATION REGIONS

5 BACKGROUND OF THE INVENTION

[0001] The present invention relates to floating gate nonvolatile memories.

[0002] A floating gate nonvolatile memory cell stores information by storing an electrical charge on its floating gate. The floating gate is capacitively coupled to the control gate. In order to write the cell, a potential difference is created between the
10 control gate and some other region, for example, the source, drain or channel region of the cell. The voltage on the control gate is capacitively coupled to the floating gate, so a potential difference appears between the floating gate and the source, drain or channel region. This potential difference is used to change the charge on the floating gate.

[0003] In order to reduce the potential difference that has to be provided between the
15 control gate and the source, drain or channel region, it is desirable to increase the capacitance between the control and floating gates relative to the capacitance between the floating gate and the source, drain or channel region. More particularly, it is desirable to increase the "gate coupling ratio" GCR defined as $CCG/(CCG + CSDC)$ where CCG is the capacitance between the control and floating gates and CSDC is the capacitance
20 between the floating gate and the source, drain or channel region. One method for increasing this ratio is to form spacers on the floating gate. See U.S. patent no. 6,200,856 issued March 13, 2001 to Chen, entitled "Method of Fabricating Self-Aligned Stacked Gate Flash Memory Cell", incorporated herein by reference. In that patent, the memory is fabricated as follows. Silicon substrate 104 (Fig. 1) is oxidized to form a pad oxide
25 layer 110. Silicon nitride 120 is formed on oxide 110 and patterned to define isolation trenches 130. Oxide 110 and substrate 104 are etched, and the trenches are formed. Dielectric 210 (Fig. 2), for example, borophosphosilicate glass, is deposited over the structure to fill the trenches, and is planarized by chemical mechanical polishing (CMP). The top surface of dielectric 210 becomes even with the top surface of nitride 120. Then
30 nitride 120 is removed (Fig. 3). Oxide 110 is also removed, and gate oxide 310 is thermally grown on substrate 104 between the isolation trenches. Doped polysilicon

layer 410.1 (Fig. 4) is deposited over the structure to fill the recessed areas between the isolation regions 210. Layer 410.1 is polished by chemical mechanical polishing so that the top surface of layer 410.1 becomes even with the top surface of dielectric 210.

5 [0004] Dielectric 210 is etched to partially expose the “edges” of polysilicon layer 410.1 (Fig. 5). Then doped polysilicon 410.2 (Fig. 6) is deposited and etched anisotropically to form spacers on the edges of polysilicon 410.1. Layers 410.1, 410.2 provide the floating gates.

10 [0005] As shown in Fig. 7, dielectric 710 (oxide/nitride/oxide) is formed on polysilicon 410.1, 410.2. Doped polysilicon layer 720 is deposited on dielectric 710 and patterned to provide the control gates.

[0006] Spacers 410.2 increase the capacitance between the floating and control gates by more than the capacitance between the floating gates and substrate 104, so the gate coupling ratio is increased.

SUMMARY

15 [0007] This section summarizes some features of the invention. Other features are described in the subsequent sections. The invention is defined by the appended claims which are incorporated into this section by reference.

20 [0008] In some embodiments of the present invention, as in prior art, the floating gate is formed from two conductive layers, but the fabrication process is different. In one embodiment, the dielectric in the substrate isolation regions (such as dielectric 210) and the first of the two conductive layers providing the floating gates (such as layer 410.1) are formed so that the dielectric has an exposed sidewall. At least the top portion of the sidewall is exposed. This can be achieved by etching down the first conductive layer to lower its top surface below the top surface of the dielectric regions. Then some of the
25 dielectric is removed from the exposed portions of the dielectric sidewalls to laterally recess the sidewalls. Then the second conductive layer for the floating gates is formed. The recessed sidewalls of the dielectric allow the second conductive layer to expand laterally, thus increasing the capacitive coupling between the floating and control gates and improving the gate coupling ratio.

30 [0009] The invention is not limited to the features described above. For example, the

floating gates can be formed from more than two layers. Other features and advantages of the invention are described below. The invention is defined by the appended claims.

BRIEF DESCRIPTION OF THE DRAWINGS

5 **[0010]** Figs. 1-7 show cross sections of prior art semiconductor memory structures in the process of fabrication.

[0011] Figs. 8-20 show vertical cross sections of memory structures during fabrication according to embodiments of the present invention.

[0012] Fig. 21 is a top view of the structure of Fig. 20.

DESCRIPTION OF SOME EMBODIMENTS

10 **[0013]** The embodiments described in this section illustrate but do not limit the invention. The invention is not limited to particular materials, process steps, conductivity types (P or N), or dimensions. The invention is defined by the appended claims.

15 **[0014]** The beginning stages of the memory fabrication can be as in Figs. 1-4. In some embodiments, the memory array is fabricated in a doped region of type P- that is formed in monocrystalline semiconductor substrate 104 (Fig. 8). This region can be isolated by P-N junctions (not shown). See e.g. U.S. patent no. 6,355,524 issued March 12, 2002 to H.T. Tuan et al. and incorporated herein by reference.

20 **[0015]** Silicon dioxide 110 (pad oxide) is formed on substrate 104 by thermal oxidation or some other technique to an exemplary thickness of 150 Å. (The dimensions are given for an exemplary process using a 0.18 µm line width technology; the line width is the minimal dimension that can be reliably printed photolithographically). Silicon nitride 120 is deposited on oxide 110. In one embodiment, the thickness of nitride 120 is in the range of 2000 Å to 2200 Å. Nitride 120 is patterned photolithographically, using a photoresist mask (not shown), to define substrate isolation trenches 130. Oxide 110 and
25 substrate 104 are etched through the openings in nitride 120. Trenches 130 (“STI trenches”) are formed in the substrate as a result. An exemplary depth of trenches 130 is 0.2 µm, measured from the top surface of the substrate 104. Other depths are possible. Trenches 130 will be filled with dielectric to provide isolation between active areas 832 of substrate 104.

[0016] In Fig. 8, the trenches have sloping sidewalls, and the trenches widen at the top. In some embodiments, the trenches have vertical sidewalls, or the trenches are wider at the bottom. The invention is not limited by the shape of the trenches.

[0017] Nitride 120 is subjected to a wet etch to recess the vertical edges of the nitride layer away from trenches 130. See Fig. 9. This step reduces the aspect ratio of the holes that will be filled with dielectric 210 (these holes are formed by the openings in nitride 120 and oxide 110 and by the trenches 130). The lower aspect ratio facilitates filling these holes. As will be seen below, the etch of nitride 120 will also reduce the capacitance between the floating gate and the substrate 104, thus increasing the gate coupling ratio.

[0018] A thin layer 210.1 of silicon dioxide is thermally grown on the exposed silicon surfaces to round the edges of trenches 130 and passivate the trench surfaces. Silicon dioxide 210.2 (Fig. 10) is deposited by a high density plasma process. Oxide 210.2 fills the trenches and initially covers the nitride 120. Oxide 210.2 is polished by a CMP process that stops on nitride 120. A planar top surface is provided.

[0019] In the subsequent figures, the layers 210.1, 210.2 are shown as a single layer 210. This dielectric 210 will be referred to as STI dielectric or, more generally, as field dielectric.

[0020] Nitride 120 is removed selectively to oxide 210. This can be done by a wet etch (e.g. with phosphoric acid). Pad oxide 110 is etched away to expose the substrate 104 in the active areas. The oxide etch may remove a small amount of oxide 210.

[0021] Gate dielectric 310 (Fig. 11) is formed on the active areas. In one example, dielectric 310 is silicon dioxide thermally grown to a thickness of 90 Å.

[0022] Polysilicon 410.1 is deposited over the wafer and is heavily doped during or after the deposition. Layer 410.1 will provide portions of the conductive floating gates. An exemplary deposition process is LPCVD. Polysilicon 410.1 fills the recesses between the STI dielectric regions 210 and covers the whole wafer.

[0023] Polysilicon 410.1 is subjected to a timed etch to lower its top surface below the top surface of dielectric 210. See Fig. 12. If the polysilicon surface is not initially planar, the polysilicon can be planarized by depositing an auxiliary layer (e.g. photoresist) having a planar top surface and adjusting the etch to obtain equal etch rates

for the polysilicon and the auxiliary layer. Other planarization methods are also possible (e.g. CMP).

5 **[0024]** In one example, the dimensions are as follows. The width W_t of each trench 130 is $0.18\text{ }\mu\text{m}$ at the top of the trench. The width W_{aa} of each active area 832 is also $0.18\text{ }\mu\text{m}$. The width W_f of each floating gate feature 410.1 is $0.16\text{ }\mu\text{m}$ (W_f is the distance between the adjacent regions 210 above the substrate 104). STI oxide 210 protrudes by an amount T_{ox} above the top surface of substrate 104. T_{ox} is between $1900\text{ }\text{\AA}$ and $2100\text{ }\text{\AA}$. Polysilicon layer 410.1 is $900\text{ }\text{\AA}$ thick. Oxide 310 is $90\text{ }\text{\AA}$ thick. Therefore, the height T_{xs} of the exposed portions of the sidewalls of oxide 210 is between $910\text{ }\text{\AA}$ and $1110\text{ }\text{\AA}$.

10 **[0025]** Oxide 210 is etched selectively to polysilicon 410.1 (Fig. 13). The etch includes a horizontal component that causes the sidewalls of oxide 210 to be laterally recessed in the direction away from the adjacent polysilicon features 410.1 and active areas 832. This can be an isotropic wet etch. In some embodiments, the top sidewall portions 210T of oxide 210 become positioned over trenches 130, while the bottom
15 sidewall portions 210B continue to overlap the active areas. In one embodiment, the isotropic etch laterally recesses the sidewall by an amount L_s in the range of $0.06\text{ }\mu\text{m}$ to $0.07\text{ }\mu\text{m}$. The etch also lowers the top surface of oxide 210 by the same amount. In addition, the etch attacks the oxide portions near the polysilicon 410.1 to form pockets 1310 in which the top surface of oxide 210 falls below the top surface of polysilicon
20 410.1. The sidewalls of polysilicon 410.1 are exposed in these pockets. In some embodiments, the sidewalls are exposed to the depth L_s measured from the top surface of polysilicon 410.1. In Fig. 13, the top surface of oxide 210 is above the top surface of polysilicon 410.1 after the oxide etch, but this is not necessary. The top surface of oxide 210 may be even with, or below, the top surface of polysilicon 410.1.

25 **[0026]** Polysilicon layer 410.2 (Fig. 14) is deposited on the structure, and is doped during or after the deposition to the same conductivity type as the layer 410.1. Layer 410.2 will provide portions of the floating gates. Layer 410.2 fills the pockets 1310. An exemplary deposition process is conformal low pressure chemical vapor deposition (LPCVD). The thickness of layer 410.2 (at least $2000\text{ }\text{\AA}$ in some embodiments) is chosen
30 to provide a planar top surface. The planar top surface is not necessary however.

[0027] Layer 410.2 is subjected to CMP or an etch to remove the polysilicon 410.2 from over the top horizontal surfaces 210H (Fig. 15) of oxide features 210 and thus to

isolate the floating gates from each other. Each floating gate will include the adjacent portions of layers 410.1, 410.2 that contact each other. In some embodiments, if the top surface of polysilicon 410.2 is not planar before the etch, a planarizing etch can be used with an auxiliary material (e.g. photoresist) as described above in connection with the
5 etch of layer 410.1. The top surface of layer 410.2 may be above, below, or even with the horizontal top surface 210H of oxide 210.

[0028] The recessed sidewalls of oxide 210 allow the floating gates to extend laterally (by the amount L_s) beyond the floating gate portions 410.1 and, possibly, beyond the active areas 832. The lateral extension of the floating gates increases the capacitive
10 coupling between the floating gates and the control gates 720 (Fig. 17) and increases the gate coupling ratio. In one embodiment, the spacing S_f between the adjacent floating gates is only 0.05-0.06 μm for the dimensions given above in connection with Fig. 13. S_f can be smaller than the minimum line width.

[0029] An optional etch of oxide 210 lowers the top surface of oxide 210 to a level
15 below the top surface of polysilicon 410.2 (Fig. 16) to increase the capacitive coupling between the floating gates 410.1, 410.2 and the control gates 720 (Fig. 17). See the aforementioned U.S. patent no. 6,355,524.

[0030] Insulating layer 710 (Fig. 17) is formed over the structure. In some embodiments, layer 710 is a sandwich of silicon dioxide, silicon nitride, silicon dioxide
20 (ONO). A conductive layer 720, e.g. doped polysilicon, is deposited over ONO 710. Layer 720 provides the control gates. Layers 720, 710, 410.1, 410.2 can be patterned as desired.

[0031] Other fabrication steps, such as doping of the source and drain regions, depend on a particular memory structure. The floating gate fabrication techniques described
25 above can be incorporated into many memory structures, known or to be invented. Some structures are shown in Figs. 18-20. These figures illustrate the memory cross-sections by a vertical plane marked A-A in Fig. 17. This plane passes through an active area 832, and is perpendicular to the cross-sectional plane of Figs. 8-17. The memory cell source/drain regions are shown at 1810. Fig. 18 illustrates a stacked gate memory cell,
30 with each control gate line 720 overlying a number of floating gates made from layers 410.1, 410.2. See U.S. patent no. 6,013,551 issued January 11, 2000 to Chen et al., incorporated herein by reference. Fig. 19 illustrates a split gate cell, in which the layer

720 provides a control gate and a select gate.

[0032] Fig. 20 illustrates a cell in which the select gate is provided by a separate conductive layer 2010. This type of cell is described in the aforementioned U.S. patent no. 6,355,524. The top view of the memory array is shown in Fig. 21. Line A-A in Fig. 21 marks the cross sectional plane of Fig. 20 (the same plane as marked in Fig. 17). The floating gates 410.1, 410.2 are shown at 410 in Fig. 21. The source/drain regions are shown as 1810S (source line) and 1810B (bit line regions connected to bitlines; the bitlines are not shown). After the deposition of polysilicon 720, a silicon nitride layer 2020 (Fig. 20) is formed on the structure. (Nitride 2020 is not shown in Fig. 21.) Then the layers 2020, 720, 710, 410.2, 410.1 are patterned using a single photolithographic mask. As a result, the layer 720 forms a number of control gate lines extending in the row direction (X direction in Fig. 21) across the memory array. Each control gate line provides the control gates for one row of the memory cells. Dielectric 2030 is formed over the sidewalls of each structure containing the layers 2020, 720, 710, 410.2, 410.1 for one row. The exposed portions of oxide 310 are etched away, and gate dielectric 2040 is formed on the exposed substrate areas for the select transistors. Doped polysilicon 2010 is deposited and patterned to form wordlines. Each wordline provides the select gates for one row.

[0033] The etch of polysilicon 410.1 and oxide 310 exposes the trench dielectric 210 and possibly the silicon substrate 104. As shown in Fig. 17, dielectric 210 overlaps the edges of active areas 832 and protects the active areas from pitting during the etch of layers 410.1 and 310. Also, the overlapping portions of dielectric 210 prevent the trench dielectric grooving at the trench edges. (The grooving, i.e. removal of the trench dielectric at the trench edges, may increase the leakage current in the active areas near the trenches.) At the same time, the top surface of dielectric 210 is recessed and in some embodiments does not overlap the active areas, providing the increased capacitance between the floating and control gates.

[0034] In another embodiment, gate oxide 310 and polysilicon 410.1 are formed before the etch of trenches 130. Then a sacrificial silicon nitride layer is deposited. The nitride layer, the layers 310, 410.1, and substrate 104 are patterned with a single mask to form the trenches. See the aforementioned U.S. patent no. 6,355,524. The trenches are filled with dielectric 210. The sacrificial nitride is removed. Dielectric 210 protrudes

upward above the polysilicon 410.1, as in Fig. 12. The fabrication proceeds as in Figs. 13-17.

[0035] The invention is not limited to the structures and methods described above. The invention is not limited to any materials or fabrication processes. For example,
5 floating gate layers 410.1, 410.2 can be formed from different materials. The top surface of layer 410.2 (Fig. 17) could be even with, or below, the top surface of layer 410.1. The invention is not limited to any memory programming or erase mechanisms. The invention includes both flash and non-flash memories. Other embodiments and variations are within the scope of the invention, as defined by the appended claims.